Dynamic ElecTronic hEalth reCord deTection (DETECT) of individuals at risk of a first episode of psychosis: a case-control development and validation study

Lars Lau Raket, Jörn Jaskolowski*, Bruce J Kinon, Jens Christian Brasen, Linus Jönsson, Allan Wehnert, Paolo Fusar-Poli*

Summary

Background Many individuals who will experience a first episode of psychosis (FEP) are not detected before occurrence, limiting the effect of preventive interventions. The combination of machine-learning methods and electronic health records (EHRs) could help address this gap.

Methods This case-control development and validation study is based on EHR data from IBM Explorys. The IBM Explorys Platform holds standardised, longitudinal, de-identified, patient-level EHR data pooled from different health-care systems with distinct EHRs. The present EHR-based studies were retrospective, matched (1:1), case-control studies compliant with RECORD, STROBE, and TRIPOD statements. The study included individuals in the IBM Explorys database who at some point between 1990 and 2018 had a diagnosis of FEP followed by schizophrenia, and psychosis-free matched control individuals from a random subsample of the full cohort. For every individual in the FEP cohort, the individual in the control cohort was matched to have a similar date for inclusion in the database and a similar total observation time. Individuals in the FEP cohort had their index date defined as the first diagnosis of psychosis or the first prescription of antipsychotic medication. Individuals in the control cohort had their index date defined to occur the same number of days after inclusion in the database as their matching FEP individual. The FEP and control cohorts were both randomly split into development and validation datasets in a ratio of 7:3. The subset of individuals in the validation dataset who had all their health-care encounters at providers that were not seen in the development dataset made up the external validation subset. A novel recurrent neural network model was developed to predict the risk of FEP 1 year before the index date by employing demographics and medical events (in the categories diagnoses, prescriptions, procedures, encounters and admissions, observations, and laboratory test results) dynamically collected in the EHR as part of clinical routine. We named the recurrent neural network Dynamic ElecTronic hEalth reCord deTection (DETECT). The main outcomes were accuracy and area under receiver operating characteristic curve (AUROC). Decision-curve analyses and dynamic patient journey plots were used to evaluate clinical usefulness.

Findings The FEP and control cohorts each comprised 72 860 individuals. 102 030 individuals (51 015 matching pairs) were randomly allocated to the development dataset and the remaining 43 690 to the validation dataset. In the validation dataset, 4770 individuals had all their encounters outside of the 118 790 health-care providers that were encountered in the development dataset. The data from these individuals made up the external validation subset. The median follow-up (observation time before index date) was 6·0 years (IQR 3·0–10·4). In the development dataset, DETECT's prognostic accuracy was 0·787 and AUROC was 0·868. In the validation dataset, DETECT's prognostic accuracy was 0·774 and AUROC was 0·856. In the external test subset, DETECT's balanced prognostic accuracy was 0·724 and AUROC was 0·799. Prevalence-adjusted decision-curve analyses suggested that DETECT was associated with a positive net benefit in two different scenarios for FEP detection.

Interpretation DETECT showed adequate prognostic accuracy to detect individuals at risk of developing a FEP in primary and secondary care. Replication and refinement in a population-based setting are needed to consolidate these findings.

Funding Lundbeck.

Lancet Digital Health 2020; 2: e229–39

Published Online March 26, 2020 https://doi.org/10.1016/S2589-7500(20)30024-8

See Comment page 210

*Contributed equally

Lundbeck, Valby, Denmark (L L Raket PhD, J C Brasen PhD, L Jönsson MD, A Wehnert MPharm); **Clinical Memory Research Unit, Lund University, Lund, Sweden** (L L Raket); **IBM Danmark, Brøndby, Denmark** (J Jaskolowski MSc); **Lundbeck US, Deerfield, IL, USA** (B J Kinon MD); **Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden** (L Jönsson); **Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK** (Prof P Fusar-Poli MD); **OASIS** (Prof P Fusar-Poli) **and National Institute for Health Research Maudsley Biomedical Research Centre** (P Fusar-Poli), **South London and Maudsley NHS Foundation Trust, London, UK** (Prof P Fusar-Poli); **and Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy** (P Fusar-Poli)

Correspondence to: Dr Lars Lau Raket, Lundbeck, 2500 Valby, Denmark **llra@lundbeck.com**

Introduction

Schizophrenia is a disabling psychiatric disorder, diagnosed in more than 21 million people around the world, with high social and economic costs for patients, their families, and society.[1–4] In patients with schizophrenia, the first episode of psychosis (FEP) is a key event that defines long-term outcomes.[5] Although antipsychotic treatments are effective,[6,7] once the first episode of psychosis occurs there are only limited possibilities to improve long-term outcomes.[8,9] Prevention or delay of

Research in context

Evidence before this study

We established the current state of risk prediction using electronic health records (EHRs) in general, and specifically for first episode of psychosis (FEP), including schizophrenia, by searching PubMed and Google Scholar for studies published in English up to Aug 19, 2019, using the keywords "electronic health record(s)", "prediction", "psychosis", and "schizophrenia". We found many recent studies that successfully applied machine-learning methods to risk prediction on EHRs, but only a few studies predicted FEP or schizophrenia. In these, the case numbers were typically low and generalisability was therefore unclear.

Added value of this study

We developed and validated Dynamic ElecTronic hEalth reCord deTection (DETECT): the first individualised risk-prediction model that leverages large-scale screening of EHRs to detect individuals at risk of FEP. DETECT was based on a recurrent neural network model, and predicted whether a FEP (which developed into schizophrenia) would occur 1 year before its onset with an area under receiver operating characteristic curve of 0·856. Prevalence-adjusted decision-curve analyses found that detection based on DETECT was associated with a positive net benefit for cost-benefit ratios as low as 1:16, suggesting potential clinical utility of the model.

Implications of all the available evidence

Early detection of most individuals at risk of psychotic disorders can improve their long-term outcomes. Successful application of machine-learning methods to detect individuals at increased risk on the basis of existing data is a low-cost approach to increase value of EHR data, with the potential to benefit patients and society. The results of our study suggest that using DETECT for large-scale screening is associated with potential clinical benefits. Future studies are needed to confirm and refine these findings in a population-based setting, and to investigate if the results can be replicated in other countries. Finally, there is a need to investigate whether EHR-based risk-prediction models like DETECT can be fully implemented in real-world clinical practice.

disease progression would therefore have a great impact on a personal and societal level.[5,9,10] However, the first rate-limiting step towards successful prevention of schizophrenia and other psychotic disorders is the detection of at-risk individuals before they experience psychotic episodes.[11] Currently, only 5–12% of individuals who will experience a first episode of psychosis are detected during their at-risk stage, before the disorder occurs.[12,13] Individuals at risk for psychosis are identified on the basis of a prototypical presentation of subtle symptoms that meet a clinical high-risk state for psychosis (CHR-P).[14–16] Identification of these symptoms is based on functional impairment[17] and help-seeking behaviours,[18] and depends on the ability of clinicians or caregivers to recognise these features and initiate a referral to specialised clinics.[19] Therefore, the detection of individuals at risk for psychosis is hampered by complex pathways to care and unsystematic approaches that lead to substantial sampling bias.[19,20] An additional limitation is that about one-third of patients with a FEP might not experience a CHR-P stage before the onset of their illness.[21,22] Finally, it is not currently possible to predict the risk of developing a FEP at the individual patient level.

See Online for appendix

One possible way of overcoming the challenges to detect individuals at risk for psychosis earlier might be to leverage electronic health records (EHRs), which are increasingly adopted across primary and secondary health-care systems.[19] EHRs represent real-word clinical information and as such they are ideally positioned to improve the detection of individuals at risk of psychosis on a large scale.[11] Furthermore, EHRs can contain large amounts of patient information in addition to symptom presentation, and thus offer a unique opportunity to drive the multimodal prediction of outcomes at the individual patient level. Finally, information contained in EHRs can be used to transdiagnostically detect individuals at risk of psychosis beyond the CHR-P paradigm.[23] Two studies using EHRs have confirmed that EHRs in secondary mental health care could improve the detection of individuals at risk of psychosis.[12,24] Our study expands this line of research by using the IBM Explorys database (one of the largest available real-world EHR databases that includes primary care and secondary care data) in combination with state-of-the-art machine-learning methods: the combination of these approaches has previously proven to be valuable in risk predictions of various diseases.[25–29] Our main hypothesis was that the combination of EHR and machine-learning methods would enable an above-chance prognostic accuracy for the personalised detection of individuals at risk of developing a FEP.

Methods

Study design and participants

This case-control development and validation study is based on EHR data from IBM Explorys. The IBM Explorys Platform holds longitudinal patient-level EHR data pooled from different health-care systems with distinct EHRs. Data were standardised and normalised using common ontologies (appendix p 2), searchable through a Health Insurance Portability and Accountability Act (HIPAA)-enabled, de-identified dataset from IBM Explorys. Individuals were seen in multiple primary and secondary health-care systems from 1990 to 2018 with a combination of data from clinical electronic medical records, health-care system outgoing bills, and adjudicated payer claims. The data span the continuum of health care and were collected and linked from over 920 000 providers, covering approximately 15% of the US population at the time of this study.[30] The database contains billions of patient-level entries in the categories of demographics, diagnoses, drug

prescriptions, procedures, encounters and admissions, observations (eg, blood pressure or body-mass index), and laboratory test results.[30] Further information on data representation and processing is available in the appendix (pp 2–3). The aggregated data that Explorys assembles from different health-care providers has been statistically de-identified to meet the requirements of 45 Code of Federal Regulations § 164.514(b). All data used from Explorys were de-identified to meet 1996 HIPAA and 2009 Health Information Technology for Economic and Clinical Health (HITECH) standards. Business affiliation agreements were in place between all participating health-care systems and Explorys regarding contribution of EHR data to the Explorys Platform and the use of these de-identified data. Patients who indicated at patient onboarding that they did not wish to have their data used for de-identified secondary use were not included in the dataset.

From the Explorys database, we identified individuals with a diagnosis of International Classification of Diseases (ICD)-9 or ICD-10 schizophrenia. For all these individuals, the outcome of interest was the index date of their schizophrenic disorder. The index date was operationalised as the date of their first diagnosis of ICD-9 or ICD-10 psychosis (including but not limited to schizophrenia, because the first diagnosis is often of a non-specific psychotic disorder) or the date of their first prescription of antipsychotic medication (figure 1). If the index date was defined by prescription of antipsychotic medication, the patient was excluded if the prescription was not followed by a diagnosis of ICD-9 or ICD-10 psychosis (including schizophrenia) within 5 years. Individuals were excluded if they had less than 1 year of data before the index date (see appendix p 2 for further exclusion criteria, psychosis ICD-9 and ICD-10 codes, and antipsychotic medications). Individuals included in our study with a FEP who later developed schizophrenia were defined as the FEP cohort.

A control cohort without psychotic diagnoses or antipsychotic prescriptions was also defined. For every patient in the FEP cohort, a random psychosis-free control patient was selected, ensuring that the matched control patient had a similar date for inclusion in the database and was observed for a similar duration (days between first and last data entry). The rationale for matching on these variables and the matching algorithm are described in the appendix (p 3). For the matched control individuals, a pseudo-index date was defined to occur the same number of days after inclusion in the database as the corresponding patient with a FEP. Importantly, the control individuals were not matched on other variables such as age or sex, since the risk-prediction model should be able to modify its predictions on the basis of differences in age and sex. Furthermore, such matching on demographic variables would hamper generalisability to the whole population.

The FEP and control cohorts included patient data from the years 1990–2018, and to be eligible for either of the cohorts a patient had to be between 5–70 years of age on the index date and have a valid record of sex.

In the general population, occurrence of a FEP is a rare event with an overall incidence of 26·6 per 100 000 person-years.[31] Because of the computational complexity of handling extremely large datasets and training of deep-learning models, a case-control design was chosen since

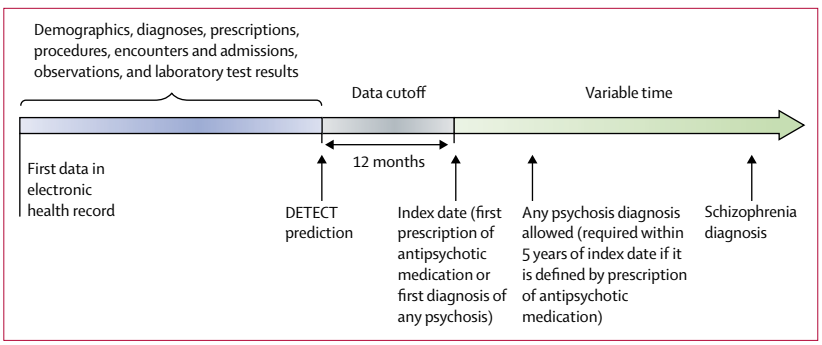

Figure 1: **Patient timeline and definition of index date for patients in the first episode of psychosis cohort**

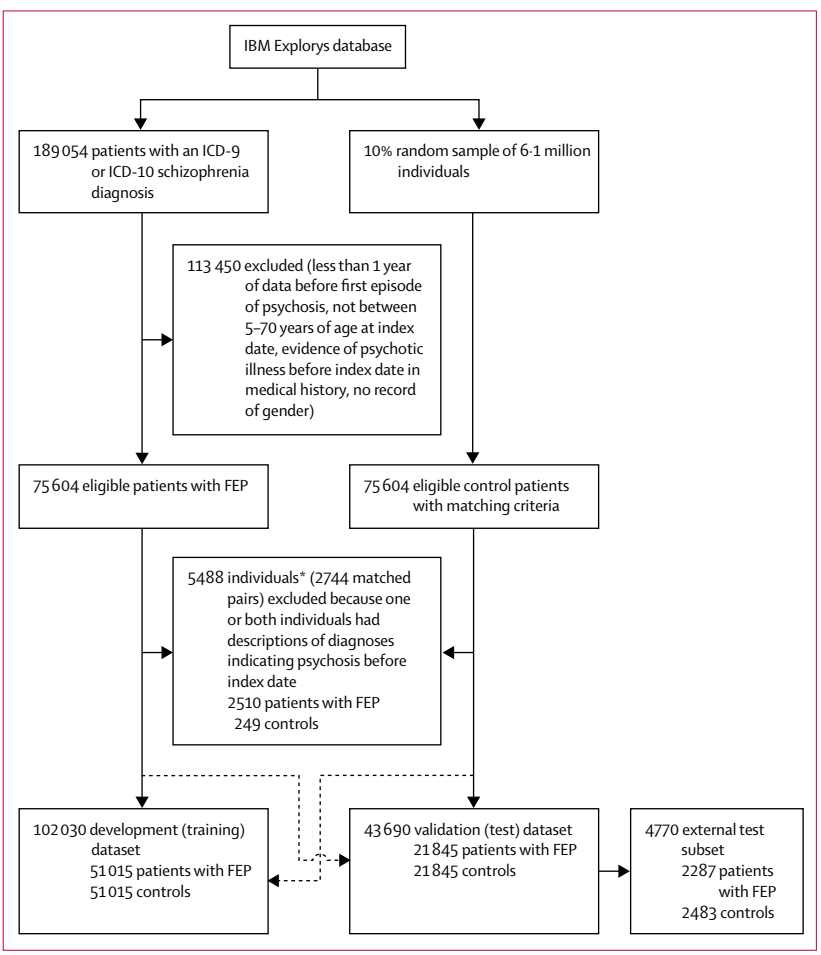

Figure 2: **Flowchart of the study population**

ICD= International Classification of Diseases. FEP=first episode of psychosis.*For 15 matching pairs both FEP and control patients have a diagnosis indicating psychosis before the index date.

	Control cohort (n=72 860)	FEP cohort (n=72 860)
Sex*		
Female	41 286 (56·7%)	31 653 (43·4%)
Male	31 574 (43·3%)	41 207 (56·6%)
Birth year†	1969 (1956–86)	1965 (1956–79)
Age at index, years†	42 (25–55)	45 (32–55)
Race*		
African American	10 347 (14·2%)	26 033 (35·7%)
Asian	1207 (1·7%)	792 (1·1%)
White	50 482 (69·3%)	38 777 (53·2%)
Hispanic or Latino	585 (0·8%)	330 (0·5%)
Multiracial	398 (0·5%)	289 (0·4%)
Other	1850 (2·5%)	1474 (2·0%)
Unknown	7991 (11·0%)	5165 (7·1%)
Ethnicity*		
Hispanic or Latino	3851 (5·3%)	3423 (4·7%)
Not Hispanic or Latino	50 010 (68·6%)	51 354 (70·5%)
Other	276 (0·4%)	123 (0·2%)
Unknown	18 723 (25·7%)	17 960 (24·7%)
Language*		
English	59 518 (81·7%)	61 666 (84·6%)
Not available	12 203 (16·7%)	10 477 (14·4%)
Other	333 (0·5%)	256 (0·4%)
Spanish	806 (1·1%)	461 (0·6%)
Insurance*		
Medicaid	2938 (4·0%)	9199 (12·6%)
Medicare	4487 (6·2%)	10 668 (14·6%)
Other	931 (1·3%)	712 (1·0%)
Other (public)	428 (0·6%)	993 (1·4%)
Private	24 200 (33·2%)	7879 (10·8%)
Self-pay	2516 (3·5%)	3833 (5·3%)
Unknown	37 360 (51·3%)	39 576 (54·3%)

Data are n (%) or median (IQR). Unknown data include those files where the patient had declined to classify. FEP=first episode of psychosis. *p<0·0001, χ^2 test. †p<0·0001, Wilcoxon test.

Table 1: **Sociodemographic characteristics of the study population**

it allowed efficient utilisation of the information for all cases in the dataset, without the computational overhead of having to train the model on the full cohort that is several orders of magnitude larger than the case group. Given the large size of both case and control cohorts and the relatively weak matching criteria, it was hypothesised that the 1:1 matched control cohort would be sufficiently representative and contain enough variety of observed patterns of medical events to allow generalisation of the findings to the full cohort.

Outcomes

The outcome of interest (an ICD-9 or ICD-10 FEP which will be associated with schizophrenia), predictors, and time to index date were automatically extracted using IBM Explorys. The recurrent neural network model would forecast the likelihood of developing a FEP 1 year before occurrence. Predictors were based on availability with no a priori selection, encompassing events that are dynamically collected from the first data entry in the EHR to the cutoff date 1 year before the index date (figure 1). Only events observed in more than 50 individuals in the development (training) dataset were included in the analyses.

Exploratory analyses of the predictions made by the recurrent neural network model were done to identify the events that had the largest effect on the predictions. Dynamic patient journey plots were generated to allow exploration of the events driving the predictions on a patient level.

Statistical analysis

This clinical register-based study is reported according to the RECORD and STROBE statements.[32] Model development and validation followed methodological guidelines[33] that have been adapted to prognostic modelling in psychiatry[34] and the TRIPOD statement.[35] For descriptive purposes, the demographic characteristics of individuals in the FEP and control cohorts were compared, and potential differences tested at a 0·05 significance level using χ^2 tests for categorical variables and Wilcoxon tests for continuous variables. The geographical distributions of individuals in the FEP and control cohorts were also compared, and state-wise differences were tested using binomial tests with a Bonferroni-corrected significance cutoff of 0·001.

FEP and control cohorts were both randomly split into development (training) and validation (test) datasets in a ratio of 7:3, respecting the pair matching (FEP-control pairs had a similar date of inclusion and follow-up time). To further validate our risk-prediction model on health-care providers that were external to the providers seen in the development dataset, unique identifiers of all health-care providers that were encountered in the development dataset were extracted, and an external test subset was defined as the subset of the validation dataset with data from patients who only had health-care encounters outside of the providers seen in the development dataset.

Model development and validation

The development dataset was used for model development and optimisation of parameters. The recurrent neural network model, called Dynamic ElecTronic hEalth reCord deTection (DETECT), was built in Keras using Python (version 3.6.7). Other analyses were done using R (version 3.3.1).[36] Briefly, the recurrent neural network model learned an embedding of medical events that mapped medical events into a geometric representation relevant for the prediction task[25] and represented an internal memory state of past events for a given patient using the gated recurrent unit architecture.[37] The model was trained to find temporal patterns of events that could separate patients in the FEP and control cohorts using data of up to 1 year before the index date. Thus, DETECT was

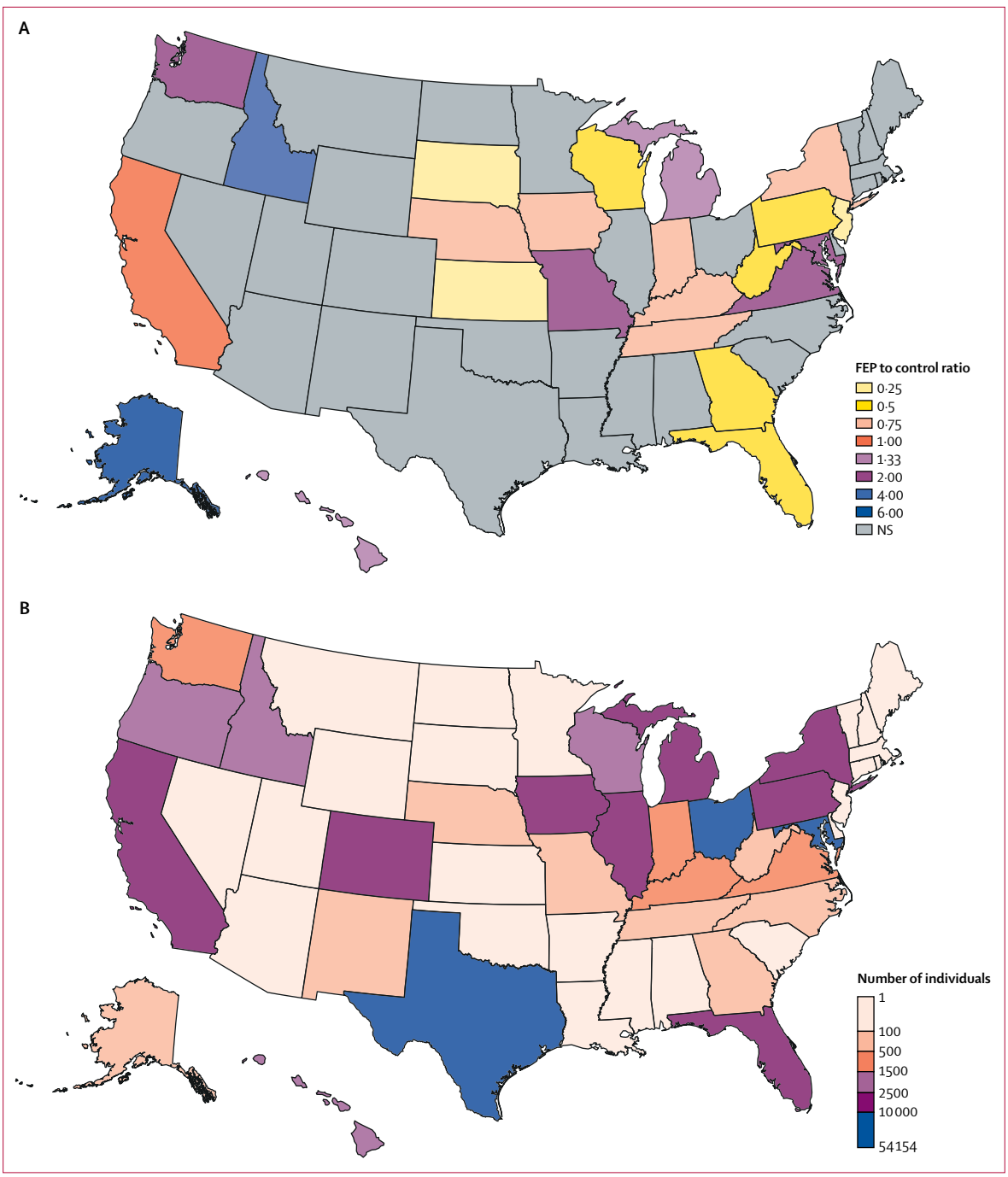

Figure 3: **Geographical distribution of individuals in FEP and control cohorts**
(A) States with significant differences in distribution of individuals in the FEP and control cohorts (state-wise binomial tests, Bonferroni-corrected significance level of p=0·001). (B) Combined geographical distribution of 145 720 individuals included in this study across US states. The cutoff is at 54154 because this was the highest number of individuals in a single state. FEP=first episode of psychosis. NS=not significant.

optimised to predict 1 year before diagnosis whether a FEP would occur. Importantly for potential prospective use in clinical routine, this learning process is dynamic because DETECT can update its prediction every time a new event enters the database. A detailed description of DETECT's architecture, parameters, and design choices is provided in the appendix (pp 3–4).

DETECT was applied to the validation dataset to measure its prognostic performance: accuracy (true classifications divided by total classifications) and area

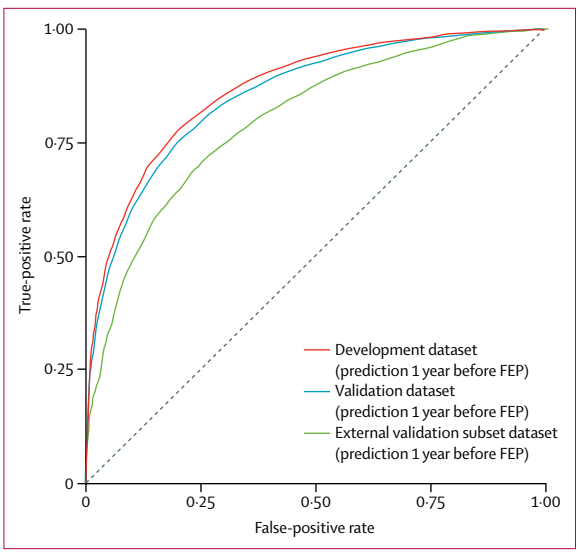

Figure 4: **Receiver operating characteristic curves**
The curves show the prognostic performance of the recurrent neural network DETECT in the development and validation datasets and in the external validation subset. DETECT=Dynamic ElecTronic hEalth reCord deTection. FEP=first episode of psychosis.

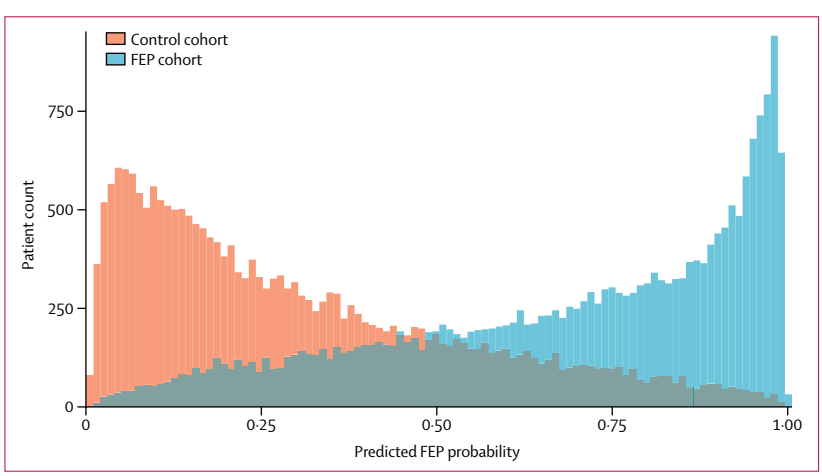

Figure 5: **Distribution of DETECT's predicted probabilities**
Distribution of predicted risk probabilities of developing a FEP 1 year before the index date, by the recurrent neural network DETECT, for the FEP and control cohorts in the validation dataset. DETECT=Dynamic ElecTronic hEalth reCord deTection. FEP=first episode of psychosis.

under receiver operating characteristic curve (AUROC), in line with methodological guidelines.[33] Since prognostic performance measures do not tell us whether DETECT would do more harm than good if used in clinical routine,[38] we additionally estimated its potential clinical utility by doing decision-curve analyses.[39] Decision-curve analyses relate the threshold probability of when an individual would be considered at sufficiently high risk to be detected to the cost of false positives (ie, detection when none is needed) and true positives. These quantities can then directly be related to the relative value (net benefit) of benefits (ie, detecting psychosis risk in EHR) and harms (unnecessary detection of psychosis risk in EHR) associated with a risk-prediction model. The net benefit values in the analysis are abstractions that simply rely on the assumption that benefits and harms are assessed using the same measurement units. In practice, benefits and harms could, for example, be assessed using quality-adjusted life-years, health-care costs, or a combination of such measures. The decision-curve analyses measured the net benefit of DETECT in two different scenarios. The first scenario considered the net benefit of a single-time prediction (eg, at a certain age for all individuals) for predicting risk for psychosis at any time in the future. The second scenario considered the net benefit of continuous-time assessment of individuals at any given time, to detect patients at high risk for developing a FEP that would lead to schizophrenia at a specific future point in time (eg, 1 year after assessment). This second scenario is more conservative since it has a more restricted target population and does not consider benefits of detecting individuals at risk who will develop schizophrenia in the longer term. To adjust for the under-sampling of controls due to the case-control design,[40] the calculations were adjusted by lifetime prevalence of non-affective psychotic disorders in a general population (first scenario)[41] and by point prevalence of schizophrenia (second scenario).[42]

To quantify DETECT's prognostic performance on data from health-care providers that did not contribute data to the development dataset, the balanced accuracy (average of true positive and true negative classifications) and AUROC were further evaluated on the external test subset.

Role of the funding source

Lundbeck funded the study, but had no role in study design, data collection, and data interpretation. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

The FEP and control cohort each comprised 72 860 individuals. 102 030 individuals (51 015 matching pairs) were randomly allocated to the development dataset and the remaining 43 690 to the validation dataset (figure 2). The median follow-up (observation time before index date) was 6·0 years (IQR 3·0–10·4). Because of the random splitting, there were no significant differences in follow-up between the development and validation datasets.

The development dataset consisted of data from encounters at 118 790 unique health-care providers. In the validation dataset, 4770 individuals (2287 in the FEP cohort and 2483 in the control cohort) had all their encounters outside of the 118 790 health-care providers that were encountered in the development dataset. The data from these individuals made up the external test subset

(figure 2). Median follow-up in the external test subset was 4·1 years (IQR 2·1–8·3). 23 390 unique medical events were observed in the development dataset, and 4899 of these occurred in more than 50 patients and were thus included as predictors.[43] There were significant differences (all p<0·0001) between the FEP and control cohorts with regards to sex, age, birth year, race and ethnicity, language, and insurance (table 1; appendix p 18). There were fewer women and white people, and more African Americans and English speakers in the FEP cohort compared with the control cohort. Also, compared with the control cohort, individuals in the FEP cohort were older, fewer had private health insurance, and a higher proportion was insured via Medicaid and Medicare (national health insurance programmes intended to help with medical costs for people with limited income and resources, older people, or people with disabilities). There were also regional differences in the distribution of individuals in the two cohorts (figure 3). Among the states with significant differences in distribution of individuals from the FEP and control cohort, Alaska stood out as having the largest over-representation of individuals in the FEP cohort compared with individuals in the control cohort (approximately 5:1), and New Jersey, South Dakota, and Kansas stood out as having the largest under-representation of individuals in the FEP cohort compared with individuals in the control cohort (all approximately 1:5).

DETECT predicted whether a FEP would occur 1 year before the index date (figure 1) with an accuracy of 0·787 and an AUROC of 0·868 in the development dataset. In the validation dataset, accuracy was 0·774 and AUROC was 0·856. In the external validation subset, balanced accuracy was 0·724 and AUROC was 0·799. The receiver operating characteristic curves are shown in figure 4. The distribution of predicted probabilities in the validation dataset is shown in figure 5. Decision-curve analysis in the validation dataset (figure 6) suggested that detection based on the predictions made by DETECT were potentially associated with a positive net benefit for cost-benefit ratios below 1:3 in the first scenario (single-point risk assessment) and below 1:16 in the second scenario (continuous-time risk assessment). Analyses on the effect of follow-up time and number of observed events on prediction accuracy and sensitivity analyses are available in the appendix (p 5).

By measuring how DETECT changed its prediction with each new recorded event, the events associated with the largest positive and negative changes in predicted probability of FEP in the validation dataset were identified. The most common of these events (more than 1000 occurrences in the validation dataset) are shown in tables 2 and 3. The top predictive events split on sex and age group are available in the appendix (pp 19–24). Across the top 100 events associated with the largest positive changes in predicted probability of FEP (appendix pp 25–27), approximately half were somatic or unspecified health-care encounters, while the remaining could be categorised as relating to psychiatric or brain health problems including substance dependence or abuse, or injury, assault, or self-harm. Illustrative examples of DETECT patient journeys showing the dynamic temporal pattern of recurrent neural network predictions for six individuals are shown in the appendix (pp 32–38).

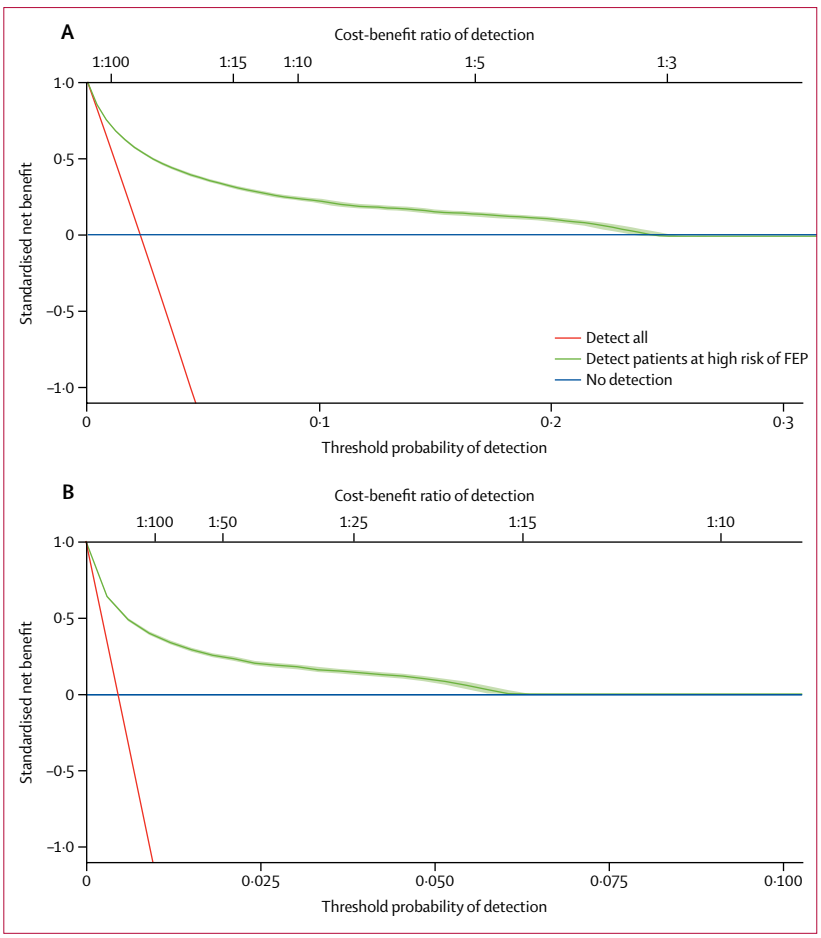

Figure 6: **Decision-curve analysis**
Decision curve analysis in the validation dataset, showing the potential clinical usefulness of DETECT in two different scenarios. Shaded bands indicate 95% CIs of net benefit estimate. (A) Single-point risk assessment for FEP. Results are for a single-time use of DETECT (eg, at a certain age for all patients) to identify individuals at high risk for ever developing a FEP. To adjust for the oversampling of cases in the case-control design, this analysis is adjusted for a lifetime prevalence of non-affective psychotic disorders of 0·0229 in the general population.[41] (B) Continuous-time risk assessment for FEP. Results are for use of DETECT to identify individuals at high risk for FEP in the immediate future. To adjust for the oversampling of cases, this analysis is adjusted for a point prevalence of schizophrenia of 0·0046.[42] DETECT=Dynamic ElecTronic hEalth reCord deTection. FEP=first episode of psychosis.

Discussion

The main aim of the current study was to develop and validate an innovative risk-prediction model (DETECT) to detect individuals at risk of developing a FEP through EHRs that contain data from both primary and secondary care. DETECT showed adequate prognostic accuracy. Further replication and implementation research are needed to consolidate these findings.

	Number of event occurrences	Ratio of events in FEP cohort versus control cohort	Median change in predicted FEP probability	
			Overall	At first occurrence
Procedure: sampling of cervix for Papanicolaou smear	2072	1·43	0·06	0·19
Procedure: differential white blood cell count	1060	14·36	0·02	0·17
Diagnosis: bipolar disorder	1777	17·51	0·03	0·15
Procedure: thin layer chromatography measurement	1083	9·72	0·05	0·11
Procedure: glucagon measurement	1226	4·55	0·02	0·11
Procedure: globulin measurement	1029	5·47	0·05	0·1
Diagnosis: seizure	1237	4·7	0·03	0·1
Procedure: valproic acid measurement	1865	27·26	0·01	0·1
Observation: normal mean blood pressure	2553	1·53	0·06	0·09
Procedure: alcohol measurement	2223	17·22	0·04	0·08
Observation: abnormally low chloride (moles/volume) in serum or plasma	2191	3·17	0·05	0·08
Diagnosis: depressive disorder	3959	2	0·05	0·07
Procedure: drug screening test	2452	6·96	0·02	0·06
Diagnosis: chronic pain	1010	2·46	0·05	0·06
Observation: abnormally low urea nitrogen (mass/volume) in serum or plasma	1583	2·69	0·05	0·06
Diagnosis: chronic obstructive lung disease	1356	2·27	0·04	0·06
Observation: normal ratio of haemoglobin A_{1c} to total haemoglobin in blood	2084	1·2	0·05	0·06
Procedure: mental state finding	1178	3·53	0·02	0·05
Diagnosis: dental consultation and report	1081	3·11	0·02	0·05
Observation: normal ratio of nucleated erythrocytes per 100 leucocytes in blood	2700	2·73	0·03	0·05
Encounter: psychiatric visit	2755	9·8	0·01	0·05
Observation: normal potassium (moles/volume) in blood	3899	1·45	0·04	0·05
Procedure: subsequent hospital visits by physician	2113	2·28	0·02	0·05
Observation: normal number of erythrocytes per volume in blood	28426	1·9	0·03	0·05
Diagnosis: long-term drug therapy	4954	1·71	0·04	0·05
Procedure: smoking cessation education	1563	0·96	0·04	0·05
Observation: normal oxygen saturation in arterial blood by pulse oximetry	4263	0·94	0·04	0·05
Observation: normal prothrombin time	1306	2·23	0·03	0·04
Diagnosis: tobacco dependence syndrome	6893	2·76	0·02	0·04
Procedure: urine pregnancy test by visual color comparison methods	3053	1·41	0·03	0·04

Top 30 events with largest positive change in predicted FEP probability after first occurrence (minimum 1000 occurrences of event in test dataset). FEP=first episode of psychosis.

Table 2: **Most predictive events in electronic health records with largest positive change**

To achieve our primary aim, we employed one of the largest real-world EHR databases. To our knowledge, this is the largest EHR study (development dataset 102 030 individuals; validation dataset 43 690 individuals; external test dataset 4770 individuals) covering approximately 15% of the US population and encompassing both primary and secondary-care data. This represents a substantial advancement given that a recent review of 91 risk-prediction modelling studies in this field found an average of 128 patients per study.[44]

DETECT was developed to reliably identify individuals potentially at risk of FEP 1 year in advance to ensure that risk prediction was done before the manifestation of more obvious behavioural pathology and functional morbidity. DETECT's prognostic balanced accuracy and AUROC in the validation dataset (0·774 and 0·856) and in the external test subset (0·724 and 0·799) were adequate, on the basis of standard performance criteria.[45] To be pragmatically useful, prognostic risk models must show above-chance performance in a population-based sample,[46] and if implemented on a large scale, even risk-prediction models with a modest accuracy could be considered of clinical utility.[47] If the results presented here are generalisable to a population-based setting, DETECT meets both of these targets, and therefore holds promise for clinical impact. To further explore the potential clinical utility of DETECT, prevalence-adjusted decision-curve analyses were done (figure 6) to illuminate the range of cost-benefit ratios for which use of DETECT was associated with net benefit in clinical use. Positive net benefits were seen for cost-benefit ratios below 1:3 in the single-point assessment and below 1:16 in the continuous-time assessment. If these results can be replicated in an external population-based setting, DETECT could potentially be useful to screen EHRs on a large scale. The primary purpose of such large-scale screening would be to detect individuals at risk of FEP who might be missed by the current health-care pathways (such as the CHR-P specialised clinics). Because DETECT uses both primary care and secondary care data it might be possible to identify individuals at risk who do not actively seek help at secondary care or CHR-P clinics and whose psychiatric risk burden might not be identified in primary care. Such individuals are not accounted for in present detection strategies for individuals at risk, because there are no risk-prediction models developed for primary care.[20] DETECT could be used in the context of a sequential and staged risk assessment framework to identify individuals at risk who might be referred to CHR-P clinics for additional face-to-face psychometric risk assessments and prognostic refinements.[48]

Another competitive advantage of DETECT is that it leverages recurrent neural network machine learning, a state-of-the-art deep learning method[49] that can mitigate other methodological biases such as overfitting and prediction in the case of infrequent events.[44] DETECT compares favourably to previous machine-learning models used to predict a FEP (or schizophrenia) from patient databases.[50,51] A previous EHR study[50] applied a stacked denoising autoencoder to data from about 700 000 individuals to develop their so-called deep-patient representation. Using this representation, a random forest algorithm predicted future diagnoses of schizophrenia in a 1-year window (occurrence between 1 and 365 days after

cut point) with an AUROC of 0·853 in a population-based internal test set (76 214 individuals). This result can largely be ascribed to the fact that the deep-patient representation included previous diagnoses of psychosis or schizophrenia and prescriptions of antipsychotic medications. DETECT had a substantially harder task since it aimed to predict a broader domain (FEP) and longer time before occurrence (365 days), without using diagnoses of previous psychoses or prescriptions of antipsychotic medication as predictors. Additionally, DETECT is innovative because it produces time-dynamic predictions that could better map the evolving phenomenological nature of emerging psychosis. At the same time, this feature of the model will allow DETECT to adapt better to future implementation in clinical routine, which is characterised by a dynamic collection of predictors. Among DETECT's predictors, we found many known risk factors[52] (table 2) and predictors that have associations with known risk factors (eg, hepatitis C and HIV infections from intravenous drug use, tests of kidney health because of lithium treatment, urinary incontinence as a side-effect of lithium or anticonvulsants). However, a substantial proportion of the most predictive events were not of a psychiatric nature (eg, sampling of cervix for Papanicolaou smear, differential white blood cell count). Although the small contributions of the majority of DETECT's predictors might seem individually irrelevant, they have a cumulative prognostic value and empirically might be associated with a more intensive use of the health-care system for individuals at risk. The patient journey plots (appendix pp 32–38) exemplify how DETECT employs the predictors, updating its prediction at every event. The patient journey plots show whether a high predicted risk is based on a few high-impact events or on many events with smaller contributions, and could help interpret the clinical significance of DETECT's predictions. This approach would mitigate the so-called black box nature of most machine-learning approaches, where the process behind predictions is opaque and eventual decisions relating to patient care are not fully transparent.[53]

This study has some limitations. First, both model development and validation were done in a case-control setting, and although minimal matching between case and control cohorts was done to ensure that the control cohort mimicked the characteristics of the full population (to improve the likelihood of generalisability of the results), the undersampling and matching of the control cohort might affect generalisability of the prediction results. Furthermore, the case-control setting might also have affected the validity of the prevalence adjustment done in the decision-curve analysis. Thus, population-based validation studies are needed before implementation can be considered. Another important limitation is that most individuals only have data covering a limited time window of their life, and data might be missing if they have used health-care providers outside of the network. Although EHR data represent data from real-world clinical practice with high ecological validity, they do not necessarily reflect research-based criteria. However, the use of structured diagnostic interviews can themselves lead to selection of patient subsamples and introduce additional bias.[54] There is meta-analytical evidence indicating that for psychotic categories, administrative data recorded in EHR are

	Number of event occurrences	Ratio of events in FEP cohort versus control cohort	Median change in predicted FEP probability	
			Overall	At first occurrence
Procedure: cytopathology procedure, preparation of smear, genital source	2331	0·28	−0·04	−0·07
Procedure: prostate specific antigen measurement	2180	0·36	−0·04	−0·07
Prescription: thyroxine	1631	0·25	−0·03	−0·06
Diagnosis: atrial fibrillation	1007	0·71	−0·04	−0·06
Procedure: surgical pathology procedure	2952	0·46	−0·05	−0·06
Procedure: review of medication	17 144	0·31	−0·03	−0·06
Diagnosis: fever	1210	0·52	−0·05	−0·06
Procedure: child health procedures	1331	0·13	−0·02	−0·06
Observation: normal prostate specific antigen (mass/volume) in serum or plasma	1569	0·49	−0·04	−0·06
Observation: abnormally low number of lymphocytes per volume in blood	1151	1·4	−0·05	−0·05
Prescription: atorvastatin	1387	0·4	−0·03	−0·05
Observation: normal bodyweight (percentile) per age	12 570	0·28	−0·02	−0·04
Diagnosis: otitis media	1435	0·35	−0·02	−0·04
Procedure: screening mammography	5204	0·3	−0·03	−0·04
Diagnosis: screening status	2467	0·31	−0·03	−0·04
Diagnosis: pre-surgery evaluation	1061	0·52	−0·03	−0·04
Prescription: metoprolol	1455	0·48	−0·03	−0·04
Procedure: *Streptococcus pneumoniae* group A antigen assay	1793	0·29	−0·02	−0·04
Procedure: comprehensive interview and evaluation	3999	0·18	−0·02	−0·04
Observation: abnormally low body-mass index (ratio)	6620	0·21	−0·01	−0·03
Procedure: weight and body-mass assessment procedure	1885	0·28	−0·02	−0·03
Observation: normal body-mass index (percentile)	9125	0·33	−0·02	−0·03
Diagnosis: coronary arteriosclerosis	1749	1·09	−0·02	−0·03
Prescription: amlodipine	1061	0·56	−0·02	−0·03
Prescription: lisinopril	2551	0·52	−0·02	−0·03
Prescription: amoxicillin	2097	0·29	−0·02	−0·03
Prescription: fentanyl	1906	0·42	−0·03	−0·03
Procedure: taking patient vital signs assessment	2064	0·21	−0·01	−0·03
Procedure: computer-assisted image analysis	1878	0·34	−0·02	−0·03
Procedure: haematology test	1564	2·65	−0·02	−0·03

Top 30 events with largest negative change in predicted FEP probability after first occurrence (minimum 1000 occurrences of event in test dataset). FEP=first episode of psychosis.

Table 3: **Most predictive events in electronic health records with largest negative change**

generally predictive of true diagnosis.[55] Furthermore, the intended use of DETECT within a sequential assessment framework mitigates these issues. Another limitation of this study is that DETECT is specifically tuned to detect a FEP in patients that will then get diagnosed with schizophrenia, and it is not tuned to detect all non-affective psychoses. Additionally, the external validation subset of data from health-care providers that did not contribute data to the development dataset was used to establish if the prediction results could be generalised to individuals who did not have any encounter with health-care providers that contributed data to the model development. Although our analysis suggested that prediction results could indeed be generalised, the population studied does not define an easily interpretable population. Since our results are geographically bound to the USA, further external (eg, geographical) validation in an independent population-based setting is needed. Crucially, the potential clinical utility of DETECT should be confirmed by pilot feasibility implementation studies that adopt a prospective real-world design. Implementation studies are scarce,[11] but should address empirical aspects of using DETECT such as ethical issues relating to the use of machine-learning methods to screen EHRs. Although patients generally welcome knowing their risk for developing serious mental disorders,[56] future participatory research with the involvement of patients (and when applicable, their carers) is required. Additional implementation research should also address the clinicians' adherence to the use of DETECT. The subsequent step would be to complete randomised large-scale studies to fully evaluate early detection of FEP using DETECT, and measure whether this model provides a clinical benefit and improves the outcomes of schizophrenia.[57–60]

Overall, our study shows that it is feasible to apply machine learning to patient-level EHR data to produce personalised, dynamic, and real-time predictions of the risk of developing a FEP. The model developed, DETECT, has adequate prognostic accuracy and holds promise of clinical utility. Before DETECT can be considered for clinical implementation it must be subjected to population-based validation, and empirical implementation challenges that might be associated with its use in clinical practice must be addressed.

Contributors

LLR and JJ developed the recurrent neural network model, collected data, developed figures and tables, and drafted the first version of the manuscript along with PF-P. All authors contributed to the study design, data interpretation, and finalisation of the article and approved the final version.

Declaration of interests

LLR, JCB, LJ, and AW are employees of Lundbeck. JJ is an employee of IBM Danmark. BJK is an employee of Lundbeck US. PF-P has served on advisory boards for Lundbeck and Menarini and has received research grant funds from Lundbeck.

Data sharing

Python code including a description of data input and trained model weights for DETECT are available under an open source licence at https://github.com/larslau/DETECT.[43]

Acknowledgments

This study was supported by Lundbeck. Medical writing support was provided by Louise Rosenberg Christ, employee of Lundbeck. Scientific support was provided by André Rausch, Associate Partner at IBM Germany.

References

1 WHO. Schizophrenia: key facts. April 9, 2018. https://www.who.int/en/news-room/fact-sheets/detail/schizophrenia (accessed Aug 19, 2019).
2 Wiersma D, Wanderling J, Dragomirecka E, et al. Social disability in schizophrenia: its development and prediction over 15 years in incidence cohorts in six European centres. *Psychol Med* 2000; **30:** 1155–67.
3 Chan SW. Global perspective of burden of family caregivers for persons with schizophrenia. *Arch Psychiatr Nurs* 2011; **25:** 339–49.
4 Almond S, Knapp M, Francois C, Toumi M, Brugha T. Relapse in schizophrenia: costs, clinical outcomes and quality of life. *Br J Psychiatry* 2004; **184:** 346–51.
5 Fusar-Poli P, McGorry PD, Kane JM. Improving outcomes of first-episode psychosis: an overview. *World Psychiatry* 2017; **16:** 251–65.
6 Zhu Y, Krause M, Huhn M, et al. Antipsychotic drugs for the acute treatment of patients with a first episode of schizophrenia: a systematic review with pairwise and network meta-analyses. *Lancet Psychiatry* 2017; **4:** 694–705.
7 Leucht S, Cipriani A, Spineli L, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *Lancet* 2013; **382:** 951–62.
8 Oliver D, Davies C, Crossland G, et al. Can we reduce the duration of untreated psychosis? A systematic review and meta-analysis of controlled interventional studies. *Schizophr Bull* 2018; **44:** 1362–72.
9 Millan MJ, Andrieux A, Bartzokis G, et al. Altering the course of schizophrenia: progress and perspectives. *Nat Rev Drug Discov* 2016; **15:** 485–515.
10 Fusar-Poli P, de Pablo GS, Correll CU, et al. Prevention of psychosis: advances in detection, prognosis and intervention. *JAMA Psychiatry* (in press).
11 Fusar-Poli P, Oliver D, Spada G, et al. Real world implementation of a transdiagnostic risk calculator for the automatic detection of individuals at risk of psychosis in clinical routine: study protocol. *Front Psychiatry* 2019; **10:** 109.
12 Fusar-Poli P, Rutigliano G, Stahl D, et al. Development and validation of a clinically based risk calculator for the transdiagnostic prediction of psychosis. *JAMA Psychiatry* 2017; **74:** 493–500.
13 Hartmann JA, Nelson B, Spooner R, et al. Broad clinical high-risk mental state (CHARMS): methodology of a cohort study validating criteria for pluripotent risk. *Early Interv Psychiatry* 2019; **13:** 379–86.
14 Fusar-Poli P, Borgwardt S, Bechdolf A, et al. The psychosis high-risk state: a comprehensive state-of-the-art review. *JAMA Psychiatry* 2013; **70:** 107–20.
15 Fusar-Poli P, Cappucciati M, Rutigliano G, et al. At risk or not at risk? A meta-analysis of the prognostic accuracy of psychometric interviews for psychosis prediction. *World Psychiatry* 2015; **14:** 322–32.
16 Fusar-Poli P. The clinical high-risk state for psychosis (CHR-P), version II. *Schizophr Bull* 2017; **43:** 44–47.
17 Fusar-Poli P, Rocchetti M, Sardella A, et al. Disorder, not just state of risk: meta-analysis of functioning and quality of life in people at high risk of psychosis. *Br J Psychiatry* 2015; **207:** 198–206.
18 Falkenberg I, Valmaggia L, Byrnes M, et al. Why are help-seeking subjects at ultra-high risk for psychosis help-seeking? *Psychiatry Res* 2015; **228:** 808–15.
19 Fusar-Poli P, Sullivan SA, Shah JL, Uhlhaas PJ. Improving the detection of individuals at clinical risk for psychosis in the community, primary and secondary care: an integrated evidence-based approach. *Front Psychiatry* 2019; **10:** 774.
20 Fusar-Poli P, Schultze-Lutter F, Cappucciati M, et al. The dark side of the moon: meta-analytical impact of recruitment strategies on risk enrichment in the clinical high risk state for psychosis. *Schizophr Bull* 2016; **42:** 732–43.
21 Rosengard RJ, Malla A, Mustafa S, et al. Association of pre-onset subthreshold psychotic symptoms with longitudinal outcomes during treatment of a first episode of psychosis. *JAMA Psychiatry* 2019; **76:** 61–70.

22 Shah JL, Crawford A, Mustafa SS, Iyer SN, Joober R, Malla AK. Is the clinical high-risk state a valid concept? Retrospective examination in a first-episode psychosis sample. *Psychiatr Serv* 2017; **68:** 1046–52.

23 Fusar-Poli P, Solmi M, Brondino N, et al. Transdiagnostic psychiatry: a systematic review. *World Psychiatry* 2019; **18:** 192–207.

24 Fusar-Poli P, Rutigliano G, Stahl D, et al. Deconstructing pretest risk enrichment to optimize prediction of psychosis in individuals at clinical high risk. *JAMA Psychiatry* 2016; **73:** 1260–67.

25 Choi E, Bahadori MT, Searles E, et al. Multi-layer representation learning for medical concepts. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, CA, USA: Association for Computing Machinery, 2016: 1495–504.

26 Lipton ZC, Kale DC, Elkan C, Wetzel R. Learning to diagnose with LSTM recurrent neural networks. *arXiv* 2015; published online Nov 11. arXiv:1511.03677 [cs.LG] (preprint).

27 Amato F, López A, Peña-Méndez EM, Vaňhara P, Hampl A, Havel J. Artificial neural networks in medical diagnosis. *J Appl Biomed* 2013; **11:** 47–58.

28 Ravizza S, Huschto T, Adamov A, et al. Predicting the early risk of chronic kidney disease in patients with diabetes using real-world data. *Nat Med* 2019; **25:** 57–59.

29 Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S, et al. Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. *JAMA Psychiatry* 2018; **75:** 1156–72.

30 IBM. The IBM Explorys Platform: liberate your healthcare data. Somers, NY, USA: IBM Watson Health, 2016. https://www.ibm.com/downloads/cas/4P0QB9JN (accessed March 10, 2020).

31 Jongsma HE, Turner C, Kirkbride JB, Jones PB. International incidence of psychotic disorders, 2002–17: a systematic review and meta-analysis. *Lancet Public Health* 2019; **4:** e229–44.

32 Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 2015; **12:** e1001885.

33 Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. Springer, 2008.

34 Fusar-Poli P, Hijazi Z, Stahl D, Steyerberg EW. The science of prognosis in psychiatry: a review. *JAMA Psychiatry* 2018; **75:** 1289–97.

35 Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015; **162:** 55–63.

36 R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2014.

37 Cho K, Merrienboer BV, Bahdanau D, Bengio Y. On the properties of neural machine translation: encoder–decoder approaches. In: Wu D, Carpuat M, Carreras X, Vecchi EM, eds. Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation. Doha, Qatar: Association for Computational Linguistics, 2014: 103–11.

38 Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* 2016; **352:** i6.

39 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006; **26:** 565–74.

40 Rousson V, Zumbrunn T. Decision curve analysis revisited: overall net benefit, relationships to ROC curve analysis, and application to case-control studies. *BMC Med Inform Decis Mak* 2011; **11:** 45.

41 Perälä J, Suvisaari J, Saarni SI, et al. Lifetime prevalence of psychotic and bipolar I disorders in a general population. *Arch Gen Psychiatry* 2007; **64:** 19–28.

42 Bhugra D. The global prevalence of schizophrenia. *PLoS Med* 2005; **2:** e151.

43 Raket LL. larslau/DETECT: DETECT (version 1.0). *Zenodo* 2020; published online March 9. DOI:10.5281/zenodo.3701836.

44 Studerus E, Ramyead A, Riecher-Rössler A. Prediction of transition to psychosis in patients with a clinical high risk for psychosis: a systematic review of methodology and reporting. *Psychol Med* 2017; **47:** 1163–78.

45 Hosmer DW, Lemeshow S, May S. Applied survival analysis: regression modeling of time to event data, 1st edn. New York, NY: John Wiley & Sons, 1999.

46 Fusar-Poli P, Werbeloff N, Rutigliano G, et al. Transdiagnostic risk calculator for the automatic detection of individuals at risk and the prediction of psychosis: second replication in an independent National Health Service trust. *Schizophr Bull* 2018; **45:** 562–70.

47 Chekroud AM, Zotti RJ, Shehzad Z, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry* 2016; **3:** 243–50.

48 Schmidt A, Cappucciati M, Radua J, et al. Improving prognostic accuracy in subjects at clinical high risk for psychosis: systematic review of predictive models and meta-analytical sequential testing simulation. *Schizophr Bull* 2017; **43:** 375–88.

49 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521:** 436–44.

50 Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Sci Rep* 2016; **6:** 26094.

51 Mechelli A, Lin A, Wood S, et al. Using clinical information to make individualized prognostic predictions in people at ultra high risk for psychosis. *Schizophr Res* 2017; **184:** 32–38.

52 Radua J, Ramella-Cravaro V, Ioannidis JPA, et al. What causes psychosis? An umbrella review of risk and protective factors. *World Psychiatry* 2018; **17:** 49–66.

53 Castelvecchi D. Can we open the black box of AI? *Nature* 2016; **538:** 20–23.

54 Webb JR, Addington J, Perkins DO, et al. Specificity of incident diagnostic outcomes in patients at clinical high risk for psychosis. *Schizophr Bull* 2015; **41:** 1066–75.

55 Davis KA, Sudlow CL, Hotopf M. Can mental health diagnoses in administrative data be used for research? A systematic review of the accuracy of routinely collected diagnoses. *BMC Psychiatry* 2016; **16:** 263.

56 Bellón JÁ, Moreno-Peral P, Moreno-Küstner B, et al. Patients' opinions about knowing their risk for depression and what to do about it. The predictD-qualitative study. *PLoS One* 2014; **9:** e92008.

57 Fusar-Poli P, Sullivan SA, Shah JL, Uhlhaas PJ. Improving the detection of individuals at clinical risk for psychosis in the community, primary and secondary care: an integrated evidence-based approach. *Front Psychiatry* 2019; **10:** 774.

58 Stafford MR, Jackson H, Mayo-Wilson E, Morrison AP, Kendall T. Early interventions to prevent psychosis: systematic review and meta-analysis. *BMJ* 2013; **346:** f185.

59 Fusar-Poli P, Borgwardt S, Bechdolf A, et al. The psychosis high-risk state: a comprehensive state-of-the-art review. *JAMA Psychiatry* 2013; **70:** 107–20.

60 Fusar-Poli P, McGorry PD, Kane JM. Improving outcomes of first-episode psychosis: an overview. *World Psychiatry* 2017; **16:** 251–65.